

7月1日-7月7日周报

周杰辉

一、科研项目情况

(1) 基本情况汇报

项目一：之江实验室-海量动态图数据的高效可视分析

工作任务：了解项目的代码框架及交互具体实现，调研 R 树与四叉树的优劣

具体进度：阅读了项目的整体代码结构，撰写文档总结调研结果。

当前面临的主要问题：代码架构目前需要有一个比较好的设计和组织，才能更好地迭代完善，否则实现起来会比较困难。

下周计划：积极参与进行代码编写和功能实现。

(2) 收获体会：要多与他人沟通交流，善于合作和总结。

二、论文阅读情况

(1) 本周论文

论文名：Top 10 algorithms in data mining^[1]

选文理由：本文介绍了数据挖掘中的十大经典算法

阅读程度：对其中的 C4.5、Aprior、EM、PageRank、kNN、Naive Bayes 和 CART 算法进行了学习

文章收获：要理解算法的思想，并能够进行代码实现。了解算法历史和演变过程并进行对比分析也十分重要，比如 C4.5 在 ID3 的基础上进行了改进，使用了信息增益比，而 CART 使用了 Gini 指数来选择特征进行分割。

(2) 下周论文计划：继续阅读上述论文，对其介绍的其余算法进行阅读学习。

三、其他事务

(1) 事务名：VAG 博客搬运

本周完成了原博客第 22-23 页的搬运，利用 Markdown 对其进行了重新排版。

(2) 事务名：小组编程训练讨论及选题

本周讨论了编程训练的工作分工，同时和胡元哲一起为小组的编程训练选了一些面向工作的类似 ACM 的竞赛试题，预计之后将在 [Virtual Judge](#) 上挂出。

(3) 事务名：制作 [Seeing Theory](#) 和 [VisuAlgo](#) 的分享 PPT

本周完成了 Seeing Theory 和 VisuAlgo 两个网站的分享 PPT，下周将在 PPT 中补充添加 [Algorithm Visualizer](#)、[USFCA-Data Structure Visualizations](#) 及 [Bokeh](#) 的相关材料。

参考文献

[1] Wu X, Kumar V, Quinlan J R, et al. Top 10 algorithms in data mining[J]. Knowledge and Information Systems, 2008, 14(1):1-37.